NPS55-79-013

NAVAL POSTGRADUATE SCHOOL

Monterey, California

DATA ANALYSIS OF THE

TEACHING AWARD BALLOTS

by

Robert R. Read

July 1979

ed for public release; distribution unlimited.

ed for:

l Postgraduate School
Monterey, CA 93940

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral T. F. Dedman
Superintendent

J. R. Borsting
Provost

This report was prepared by:

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER NPS55-79-013	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Data Analysis of the Teaching Award Ballots		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Robert R. Read		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, Ca. 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, Ca. 93940		12. REPORT DATE July 1979
		13. NUMBER OF PAGES 57
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Selection Procedures
Non-Parametric Selection Methods
Ballot Data

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The report studies the data and procedures used to select the recipient of the Award for Excellence in Teaching. Several biasing features in the data summaries and record keeping are uncovered. A new scoring system is proposed and an improved data processing system is detailed. A data monitoring system is proposed which includes sensitivity analysis of the scoring as the values of arbitrary input quantities are varied.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

DATA ANALYSIS OF THE

TEACHING AWARD BALLOTS

by

Robert R. Read

I. INTRODUCTION

The annual award for excellence in teaching was established by the Superintendent in 1969. The first recipient was honored at the Spring graduation in June 1970. There is a cash prize (which has become substantial due to an endowment by Rear Admiral John J. Schieffelin) and, in recent years, the committee has provided the Provost with a list of honorable mention faculty who, as a result, have received step increments in pay. Further, the winning instructor's name is etched on a plaque in the library.

The recipient of the award is chosen by a committee of faculty which conducts a secret poll of on board students and recent former students. The data produced by the poll are machine processed. The output is coded so that the numerical summaries cannot be associated with the names of the particular faculty involved, and the winner is selected in a totally objective fashion. The membership of the committee rotates. Each new committee has exercised its responsibilities in terms of defining the set of faculty eligible to receive the award and the set of franchised voters to participate in the poll. Also, the specific use made of the numerical summaries may change from committee to committee. Major changes in the structure of the ballot have been avoided since such changes would entail substantial changes in the computer codes.

The basic format of the ballot was set by the initial committee as was the nature of the data summaries, and the author played a substantial role in these activities. Over the years, the uses made of the balloting data have increased (i.e. an honorable mention list is extracted and previous year's performance data is used in selecting the winner) and the time came for revising the data processing effort in the light of this. Because of this need the author was placed again on the selection committee (1978) and, after study, recommended revisions in processing, monitoring, and record keeping. The purpose of the present report is to document the support for and nature of the changes.

The basic recommendations are fourfold:

i) Modify the basic scoring system so that

 (a) the correlation between high scoring instructors and instructors who are identified on few ballots is removed;

 (b) the practice of listing large numbers of comparable faculty on a ballot in an effort to boost the prime candidate is discouraged.

ii) Prepare paired comparison studies in parallel to the basic scoring system so that attention is drawn to those instructors whose basic scores are relatively low because the set of instructors with whom they were compared was an unusually strong set.

iii) Organize the record keeping process so that there is a one-to-one correspondence between instructors and their identification numbers. Maintain historical records of each instructor's performance in previous years' balloting coded by A, B, E, I, representing, respectively, top 5%, next 15%, eligible but not in top 20%, ineligible.

iv) Monitor the data each year so that the goals in i) and ii) are maintained. Perform ballot population studies so that the selection committee will be made aware of anomalies in voting patterns.

The organization of the report is as follows. Section II contains some historical information about the structure of the ballot, control of errors, and choice of notation. In Section III is described the early approach to summarizing the data, scoring, and choosing the winner. The pertinent experiences and biases observed in the data processing are described in Section IV. A paired comparison study of the data is given in Section V. This leads to Section VI which summarizes the results and describes the newly implemented system. Four appendices are included which contain many of the supporting details and data analyses.

II. BACKGROUND AND NOTATION

The first Selection Committee (chaired by Prof. T. Gawain) struggled with the issues of what data to collect and how to collect it. It decided to gather information, by secret ballot, largely from the consumers of instruction. Originally, the set of franchised voters consisted of students on board, faculty, curricula officers, and recent alumni. Each voter was asked to specify the population of instructors (from a list of eligible instructors) that they felt competent to compare, and from this population they were asked to name their first, second, and third choices. The ballot had to list at least five instructors in order to qualify for inclusion in the system. Also some biographics were collected such as voter category (student, alumni, faculty, curricular officer) and curricular area (aeronautical engineering, operations research, etc.). Each ballot contains space for voluntary written remarks supporting its first choice.

The ballots are reviewed for general validity by a neutral source and the data are transferred to IBM punched cards. Each ballot is given an index number so that it may be retrieved if necessary. The written comments are coded for the computer only in terms of their presence or absence. The eligible instructors are coded with a four digit ID number. The first three digits are an index and the last is a parity check used to catch keypunch errors. The parity check involves matching this digit with the last digit of the dot product of the first three digits in the ID with the vector (3,5,1).

It is useful to think of the returned ballots as a huge data matrix. The rows correspond to the eligible faculty and each column represents a ballot. Thus each column contains at least five entries with a one opposite the row matching the faculty member receiving the first place vote, a two in the row representing the second choice, a three in the row representing the third choice, and some mark in all positions corresponding to other faculty identified by that ballot. All other positions remain void.

At the right border of this matrix one can tally four quantities X_1, X_2, X_3, N for each instructor. For $i = 1,2,3$ the value X_i is the number of votes of rank i received, and N is the number of ballots that identify the particular instructor. These four quantities have been used to create a score of the form

$$S = (w_1X_1 + w_2X_2 + w_3X_3)/N \qquad (2.1)$$

where w_1, w_2, w_3 are nonnegative weights attached to the values of first, second, and third place votes, respectively. The early computer programs use the numbers 4, 2, 1 for the weights. Also, for each instructor, is recorded the number D of ballots that contain a statement supporting his nomination for first place. Clearly $D \leq X_1$ for each instructor.

At the bottom border of this data matrix, one can tally the number K of eligible faculty identified on each ballot. These numbers were used for monitoring purposes to see whether the voters in the various categories and curricular areas had the same distributions for number of instructors identified per ballot. Further use will be described later.

It is useful to comment on the extreme inbalance of the data. If all ballots identified the same number of instructors and all instructors were identified by the same number of ballots then a highly defensible selection procedure could be devised. In reality we are a long way from this ideal. Some arbitrary choices must be made and the performance of the system must be monitored.

III. EARLY APPROACH TO BALLOT DATA ANALYSIS

The score S was used to rank the instructors in decreasing order. This score did not necessarily determine the winner. It was used primarily as a device to order the instructors so that the strongest ones appeared together at the top of the list.

The score formula (2.1) carries some presumptions with it. Division by N presumes that a good faculty member's ability to poll ranking votes, i.e. the X_i, is proportional to N the number of ballots that identify him. The value N, in turn, is thought to be roughly proportional to the number of students he has taught. Thus division by N was viewed as an equilizer for the problem of unequal exposure to students by faculty. The system of weights represents the worth of first, second, and third place votes relative to each other. The choice (4,2,1) was made arbitrarily and not without considerable dissent. Several alternative weighting systems were proposed including the set (3,2,1) which is the one offered by Condorcet, Laplace, and others in theories of elections (see [1,4]). Note that only the relative proportions of the weights are important, not their individual values. Clearly the set should satisfy the constraint

$$w_1 \geq w_2 \geq w_3 \geq 0 \quad . \tag{3.1}$$

The set (3.1) will be called the set of admissible weights.

Historically, the first selection committee agreed rather quickly that it was appropriate to divide each X_i by N. It had difficulty in deciding what to do next. At this point the data summary consisted of a vector (for each instructor)

$$x = (x_1, x_2, x_3) \tag{3.2}$$

in the first octant of three space, where $x_i = X_i \div N$. The goal is to choose one of these points as the 'best' one in an agreeable and defensible way. Generally there are two or three hundred of these points.

The following screening procedure was applied and successfully reduced the number of points under consideration to a handful. The idea (borrowed from game theory) is to eliminate from consideration all eligible faculty who are "dominated." An instructor's score vector x' is said to be dominated if there exists another score vector x'' such that the projected (one-dimensional) scores using Equation (2.1) satisfy

$$S' < S'' \quad \text{for all admissible weights.} \tag{3.3}$$

<u>HISTORICAL NOTE</u>: The original computer codes performed this screening in two stages. The first stage searched for vectors x'' that dominated x' "absolutely." That is

$$x_1' \leq x_1'', \qquad x_2' \leq x_2'', \qquad x_3' \leq x_3''$$

with strict inequality in at least one of the three positions. The second stage considered only the remaining score vectors and eliminated those that were "completely dominated" according to (3.3). This is readily accomplished as the set of admissible weights (normalized so that $w_1 + w_2 + w_3 = 1$) is a convex set having extreme points (1,0,0), (1/2, 1/2, 0), and (1/3, 1/3, 1/3). Since a convex set is determined by its extreme points, and because of the separating hyperplane theorem, the inequality (3.3) need only be checked at these three points. If $S' < S''$ at all three points then S' is completely dominated and removed from further consideraation. The committee can restrict its attention to the undominated instructors.

The data from the first year of the award are interesting. The screening by absolute dominance resulted in thirteen eligible teachers from an original list of 249. Further screening using complete dominance reduced this to three.

The three x vectors and corresponding values of N appear below.

		N	x_1	x_2	x_3
	T_1	58	.483	.172	.086
Instructor	T_2	59	.508	.136	.051
	T_3	106	.509	.123	.047

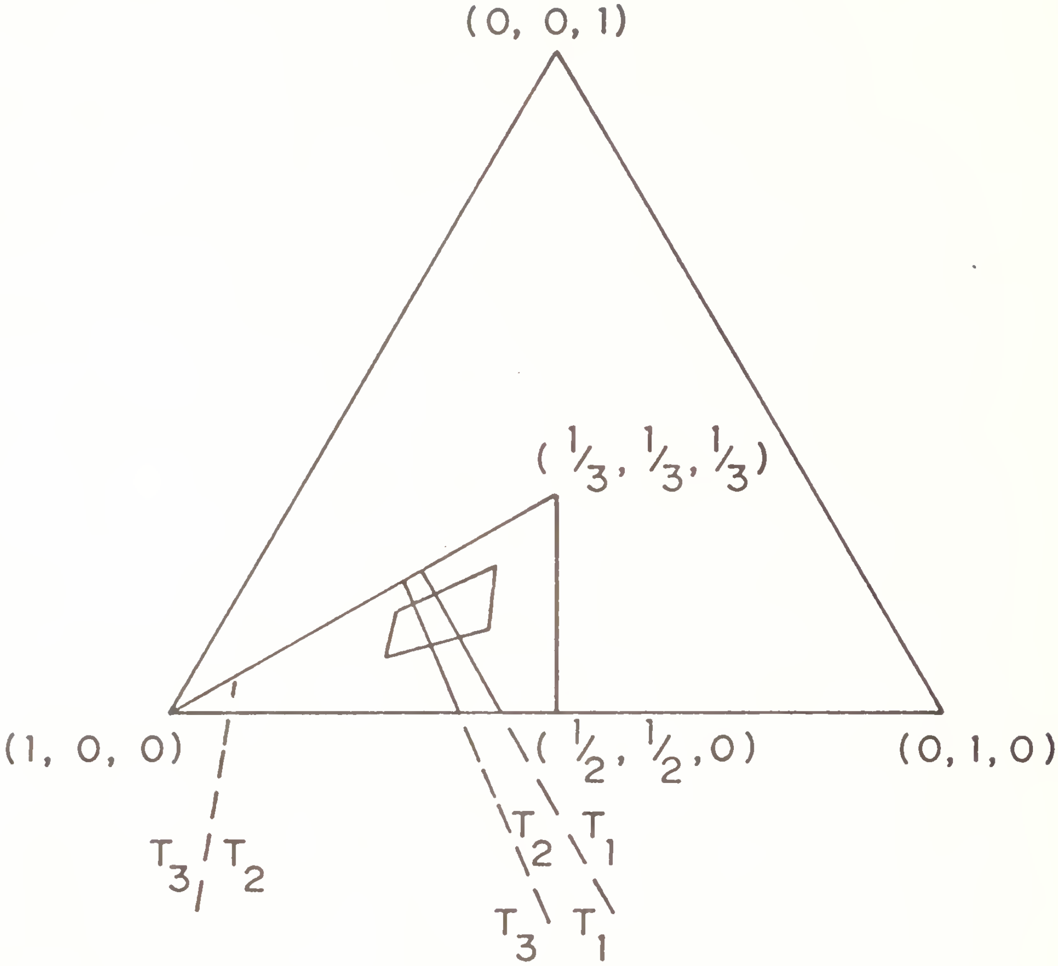


Figure 3.1

PARTITION OF THE SET OF ADMISSIBLE WEIGHTS

The subscripts of the instructors represent ranks according to the initial ordering. If the weights are allowed to "float" then each of the three remaining instructors can achieve the highest score depending upon how the weights are chosen. The choices are presented in Figure 3.1 using barycentric coordinates ([10]). The coordinate system is the simplex (w_1, w_2, w_3) constrained by $w_1 + w_2 + w_3 = 1$ and all $w_i \geq 0$. The smaller right triangle is the set of admissible weights (3.1). The three line segments serve to illustrate the weights that produce the equality of the pairs of scores $S_3 = S_2$, $S_3 = S_1$, and $S_2 = S_1$. The two outside lines partition the right triangle into three subsets which determine the decision rule. If the weight vector is chosen in the left subset then T_3 is the winner. It it is chosen in the right subset then T_1 is the winner, and otherwise T_2 is the winner. It is noted that line $S_3 = S_1$ does not affect the decision partition because T_i can win only if $S_i > S_j$ for both values of $j \neq i$. For reference purposes a quadrilateral has been drawn depicting the region where the relative weights satisfy the inequalities

$$1/3 \leq w_2/w_1 \leq 2/3, \quad 1/3 \leq w_3/w_2 \leq 2/3 \quad .$$

The weighting scheme having the ratio 4:2:1 is at the center of this quadrilateral.

By using this screening procedure the committee can reduce its choices to a few (in this case, three) and make its selection without directly choosing a set of weights.

It can also seek further information about the choices. For example the quality of the supporting statements can be (and has been) examined.

Returning to the data of the first year of the award, a modest "paired comparison" analysis provided the selection committee with some additional information about the instructors T_1, T_2, and T_3. There was only one ballot that identified both T_1 and T_2 and it made no preference. There were 19 ballots that identified T_1 in conjunction with T_2 and 13 other ballots that identified both T_2 and T_3. These "head-to-head" comparisons provided the information that generally T_2 was preferred to T_3 by the 13 ballots and that T_3 was preferred to T_1 by the 19 ballots. Some additional support for T_2 was found by including some information contained in further paired comparisons among the "top twenty" teachers as defined by the score (2.1) using 4:2:1 weights. An arbitrary rule was adopted that allowed direct comparisons only if at least five ballots identified both members of the pair. (One exception was allowed in the case of the three ballots that unanimously preferred T_3 to T_4.) The result was a set of "strings" ("$\sim$" denotes "tied score, non-zero" and ">" denotes "is preferred to")

$$T_2 > T_{15} \sim T_3 ,$$

$$T_2 \sim T_5 > T_3 > T_4 > T_{11} > T_1, \quad \text{etc.}$$

with virtually no conflicting preference patterns.

This information was made available to the selection committee for use as they saw fit. The complete data for these paired comparisons is contained in Ref. [7]. The vote of the selection committee resulted in the choice of T_2. Only at this point was the winner's name unveiled.

The data were studied for other characteristics (Ref. [7,2]) and a few are mentioned now. A principal component analysis of the x_1, x_2, x_3, revealed that the direction (in weight space w) of maximum variance is very nearly 4:2:1. Using these weights, the distribution of positive scores is exponential, specifically

$$P\{S > z | S > 0\} = e^{-2z}$$

The distributions of N (number of ballots that identify particular faculty) and K (number of faculty identified on the ballots), are both skewed positively. For 1970, the means and standard deviation are

$$\mu(N) = 64.5 \qquad \sigma(N) = 40.4$$

$$\mu(K) = 15.5 \qquad \sigma(K) = 10.5$$

About 50% of the on-board students returned their ballots. Chi-square contengency table analyses showed no evidence of dependence of score on either voter category, or on the instructor's academic rank.

IV. EXPERIENCE, TRANSITION AND RECENT STUDIES

In each of the first four years that this system was in use a master's thesis was written [2,3,7,11] which served to monitor the process and look for anomalies. A major anomaly that was noticed early was that many faculty voters tended to identify excessively large numbers of other faculty on their ballots. This was viewed as disingenuous and, by 1975, resulted in the disfranchisement of faculty as voters. At about the same time it was decided to invalidate ballots which identified too many faculty. Specifically, the voters were instructed to identify a number K of eligible faculty in the range

$$5 \leq K \leq 25 \tag{4.1}$$

Also, since the curricular officers usually are also alumni, they were removed from the voter category list.

The selection committees of 1973-74 recognized that for each winner there were a substantial number of others equally deserving of recognition. They recommended that multiple awards be made, but this was rejected by the administration. Instead, the custom began of identifying a set of ten or so "honorable mentions" and these faculty have always been recognized with pay step increases.

As more data and experience become available, the selection committee began to include "past performance" data in making their selection. In recent years there has been a very strong tendency to choose the winner from among those high scoring faculty who also have a good historical record in the honorable mention list. This practice has introduced some rather unfortunate biases since the record keeping process did not keep track of the corresponding historical eligibility of faculty. Moreover it was noticed that the honorable mention list had a bias favoring faculty with small values of N (number of ballots identifying). Such a bias leads to a tendency of favoring those who do not teach many students.

In order to respond to these problems it was necessary to perform some exploratory data analysis to complete ballot data. Unfortunately no prior year's data had been saved. The author undertook this work at the close of the 1978 balloting and those data as well as the current 1979 data were studied. The important things learned are described next.

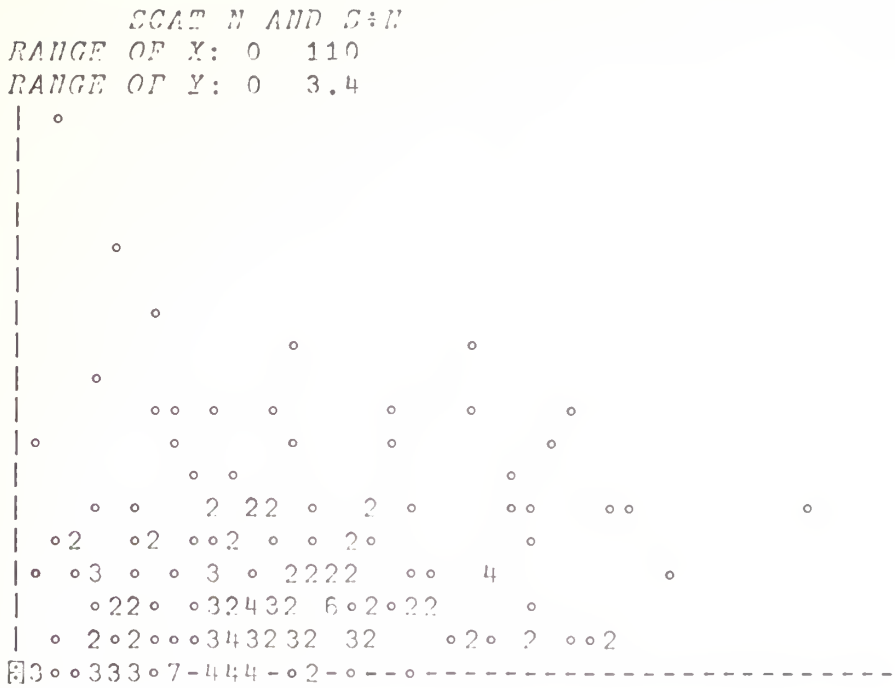
Although the selected winner is not necessarily the individual with the largest of the aforementioned scores, S, the top ten have been determined by this score and the arbitrariness of this score has come under scrutiny. Moreover, as already stated, this scoring system tends to favor faculty having the smaller values of N. This is illustrated in Figure 4.1 which contains scatter plots of score (using 4:2:1 weights) against N. The top diagram shows the bias favoring instructors with small N. The bottom diagram shows how the scatter plot would look if no normalization by N were applied. Thus, under the current system, the less well-known instructors have the advantage.

Some studies have been made in an effort to correct this situation. First, a modified score of the form

$$S' = (w_1X_1 + w_2X_2 + w_3X_3)/N^p \qquad \text{for } 0 \leq p \leq 1 \qquad (4.1)$$

was considered for several values of the exponent p and several sets of weights. According to this model a faculty member's ability to accumulate ranking votes is an increasing concave function of N. The effect is illustrated in Figure 4.2 which contains scatter plots of S' vs N^p for values of $p = .75, .5, .25$ (and the same 4:2:1 weights). Interestingly, it turns out that the exponent $p = .75$ does a good job of making the high scores uncorrelated with N and this result appears to hold for a rather large spectrum of weights. Support for this point is contained in Appendix A.

Deeper considerations uncovered the fact that the ballots contain more information than was being used. Specifically it was noticed that under the present scoring system a first place vote (or any other rank for that matter) counts the same regardless of the number of faculty with whom he is compared, i.e. it is the same whether that individual is the best of five or the best of twenty-five. Clearly the latter condition provides more



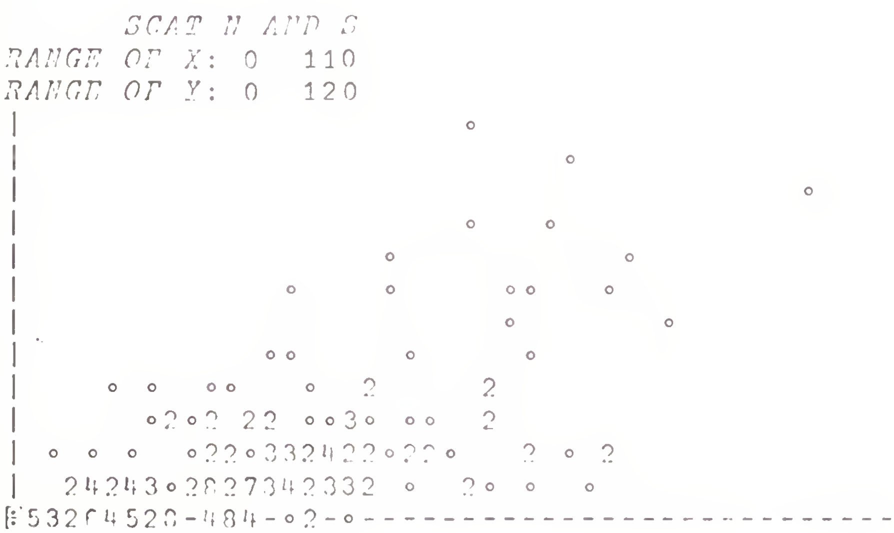


FIGURE 4.1

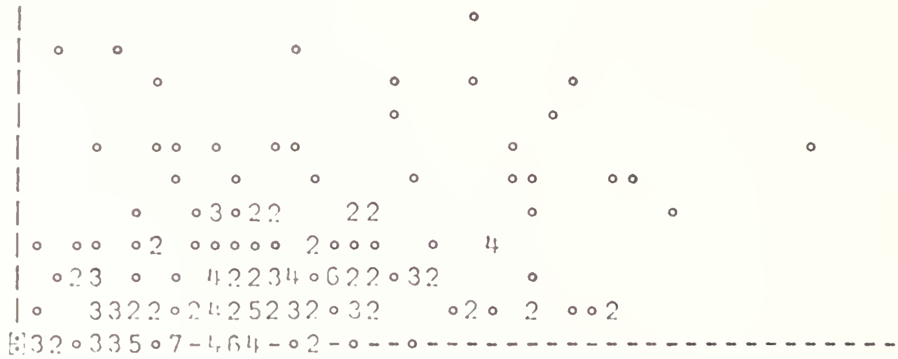
Scatter Plots of N vs Score for p = 1,0

1978 Data

SCAT N AND S SC .75

RANGE OF X: 0 110

RANGE OF Y: 0 1



SCAT N AND S SC .5

RANGE OF X: 0 110

RANGE OF Y: 0 1

SCAT N AND S SC .25

RANGE OF X: 0 110

RANGE OF Y: 0 1

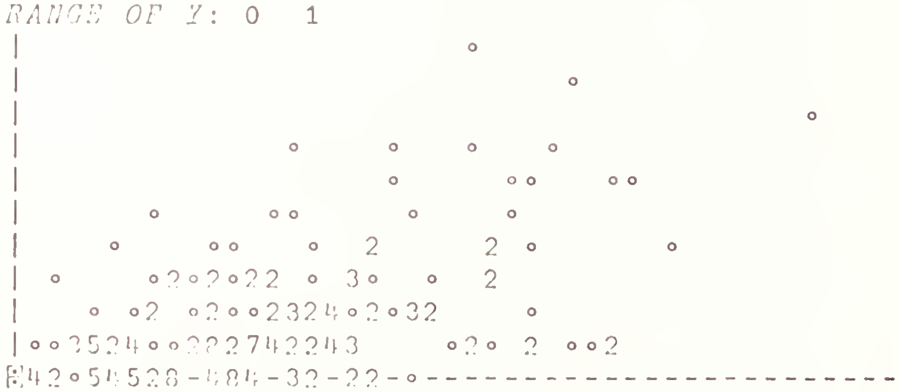


FIGURE 4.2

Scatter Plots of N vs Score for p = .75, .5, .25

1978 Data

information than the former and it is proposed that this be reflected in the accounting system. Specifically let us view each ballot as a rank distribution. If a ballot identifies a total of K instructors, then the top ranked individual on that ballot represents the $K/(K+1)$th quantile of that distribution, the second ranked individual represents the $(K-1)/(K+1)$th quantile, and the third the $(K-2)/(K+1)$th quantile. All others must be viewed as tied for positions 4 through K and they are awarded the average rank, i.e., the $(K-2)/2(K+1)$th quantile. Using this system each ballot can contribute more information to the data summary and to the score of each of its identified instructors.

More specifically, the new information collected for each instruct~~ion~~or consists of N (as before) and

$$
\begin{aligned}
Z_1 &= \sum K/(K+1), \\
Z_2 &= \sum (K-1)/(K+1) \\
Z_3 &= \sum (K-2)/(K+1) \\
Z_4 &= \sum (K-2)/2(K+1)
\end{aligned}
\tag{4.2}
$$

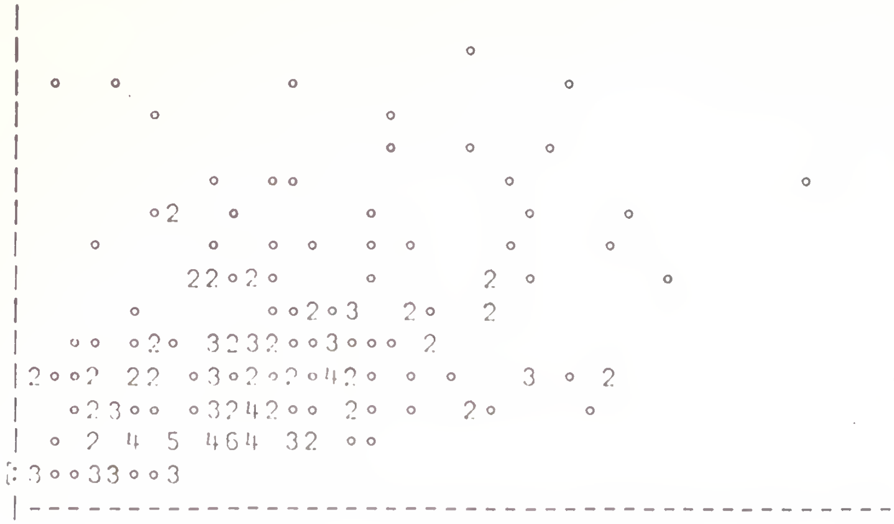
where the sums are taken over all ballots that identify that instructor. With this change, a new weighted scoring system is proposed, viz.,

$$S^* = (w_1 Z_1 + w_2 Z_2 + w_3 Z_3 + w_4 Z_4)/N^p \tag{4.3}$$

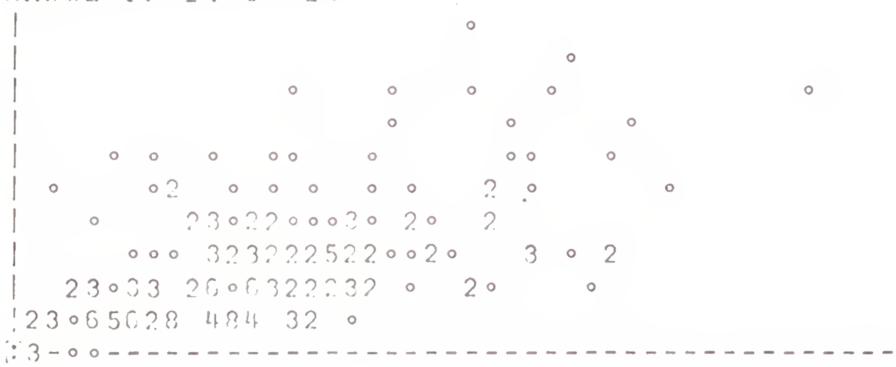
The behavior of the scores S^* was studied (using the 1978 data) for the weighting systems (3:2:1:0), (5:4:3:1), (4:3:2:1), (1:1:1:0), and (1:1:1:1) and $p = .75$, .5 and .25. Figure 4.3 contains the scatter plots of S^* vs N^p for the 5:4:3:1 weights. Again $p = .75$ produces scores whose large values are uncorrelated with N. Also this result appears to be robust as the weighting system is allowed to vary over the above listed systems. Appendices A and B contain further data supporting these points.

Although the scatter plots exhibit shape that does not depend on the weights, the system of weights is quite important in determining exactly who becomes an "honorable mention" and exactly what is placed in the "past performance" record of faculty. Since the newly proposed score has these arbitrary inputs it is important to monitor its behavior over time and adjust the inputs if necessary so that the high scores remain uncorrelated with N. Further, since the inputs are arbitrary so are the resulting rankings of faculty, and the past performance record should not contain either score or rank but a general index or grade of performance. Since the system is geared to identify good instructors it is recommended that a number, say 5%, be designated as class A instructors and the next layer, say 15%, be remembered as class B instructors. The remaining eligible faculty should be so marked and the values of N should be recorded for all. The question of how to designate the class A and B instructors is discussed later.

SCAT N AND S÷N*.75
RANGE OF X: 0 110
RANGE OF Y: 0 7.5



SCAT N AND S÷N*.5
RANGE OF X: 0 110
RANGE OF Y: 0 20



SCAT N AND S÷N*.25
RANGE OF X: 0 110
RANGE OF Y: 0 55

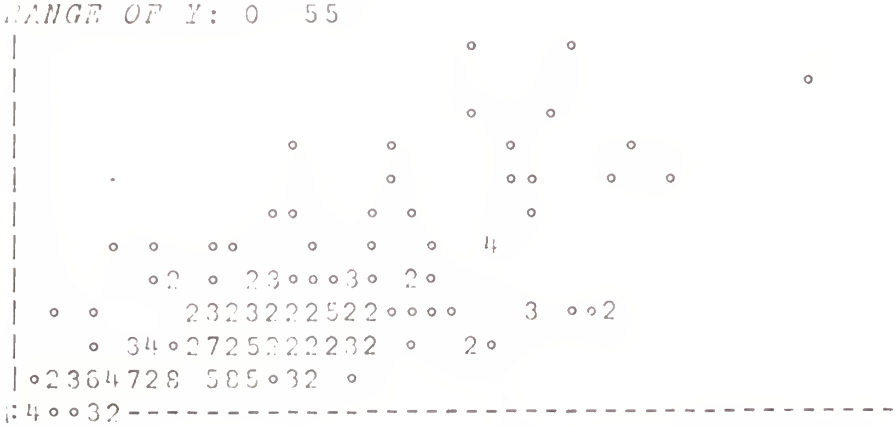


FIGURE 4.3

Scatter Plots of N vs Proposed Scores for p = .75,.5,.25.
1978 Data

V. PAIRED COMPARISONS

The proposed scoring system utilizes more, but not all, of the information contained in the ballots. Since each ballot makes some direct comparisons between individual faculty it should be possible to gather and summarize this direct information. The construction of a square win-loss matrix is useful to this end. The order of this matrix is the number of eligible faculty. The entry in row i and column j is the number of ballots that rate instructor i higher than instructor j. Ties (i.e. both i and j identified on a ballot but neither awarded a ranking position) count one-half each. All diagonal $(i = j)$ entries are zero.

The first step in analyzing a win-loss matrix is to identify the subsets of comparable instructors. Thus the list of instructors must be partitioned into several sets. If two instructors (i,j) appear in the same set then there exists a string (possibly empty) of instructor indices $(k_1, \ldots , k_r)$ such that all consecutive pairs in the string $i, k_1, k_2, \ldots, k_r, j$ are directly comparable (i.e., either the (k_s, k_{s+1}) entry in the win-loss matrix or its transpose (k_{s+1}, k_s) or both are positive). If two instructors appear in different sets then they are not directly comparable themselves, and there exists no intermediate string of pairwise directly comparable instructors connecting them. In this way the class of instructors is partitioned into sets. Within each set there is information that has a direct bearing on the question of ordering the instructors in that set. For instructors belonging to distinct sets, there is no information for comparing them.

Scaling methods are available to order linearly all instructors within each set. This does not mean that such orderings within sets are unique. If the original data are highly coherent (i.e. little controversy among the voters in forming direct comparisons) then different ordering methods will produce essentially the same results. Different methods may not produce the same orderings within a subset if there is substantial disagreement in the direct comparisons provided by the voters. The degree of coherence (or disagreement) within an individual subset can be indicated in two ways: 1) Compute a 'stress' function that measures the comparative difficulty in assigning positions to instructors, and 2) Perform the linear ordering in more than one way and compute a measure of correlation or concordance of the results. The remainder of this section deals with the description of two methods for ordering the competitors in a win-loss array and a comparison of how they perform when applied to our ballot data.

Classical Scaling

If A is a win-loss matrix then the (i,j)th entry of $N = A + A^T$ (A transpose) is the total number of contests between i and j. If A is divided by N elementwise, the i,jth element of the resultant, p_{ij}, is the empirical probability that i is preferred to j (according to all those who have examined both). This probability is converted to an approximately normal random variable, θ_{ij}, by use of the Freeman-Tukey transformation [6].

The classical scaling model Ref. [8] assumes that the positions of the teachers, T_1, T_2,... are interval scale quantities on a normal scale. Thus each θ_{ij} is viewed as an estimate (or realization) of $T_i - T_j$. The scale values T_1, T_2,... are estimated by means of a weighted least squares, i.e.,

$$\min_{T_1,\ldots,T_k} \sum_{i\neq j}\sum w_{ij}(\theta_{ij} - T_i + T_j)^2 \tag{5.1}$$

where the weights

$$w_{ij} = N_{ij} + .5 \tag{5.2}$$

are inverse variances of the θ_{ij}. A unique solution is available for each subset of comparable instructors. The partial derivatives with respect to T_r, when set equal to zero, lead to the system of equations

$$T_r \sum_j w_{rj} = \sum_j w_{rj}(\theta_{rj} + T_j) \tag{5.3}$$

which can be solved iteratively, initializing all $T_j = 0$. Let $\{\hat{T}_i\}$ be the solution and define

$$\phi_i = \sum_{j \neq i} w_{ij}(\theta_{ij} - \hat{T}_i + \hat{T}_j)^2 \tag{5.4}$$

as the stress value for T_i, i.e., it is a measure of the difficulty in fitting T_i into a linear scale containing the others. Note that the total stress $\sum \phi_i$ is the originally minimized objective function. If it is near zero, then the data are well behaved. If it is large then the data are being forced, under considerable stress, into a line. The quantity

$$\{\sum \phi_i / \sum \sum w_{ij}\}^{1/2} \tag{5.5}$$

is a pooled standard deviation of the residuals of the scaling intervals $\hat{T}_i - \hat{T}_j$.

Ford Weights

The model of Lester Ford, Jr., Ref. [5], assumes the existence of a system of weights or "odds" $\{w_i\}$ having the property that, p_{ij}, the probability that i is preferred to j, is given by

$$p_{ij} = w_i/(w_i + w_j) \tag{5.6}$$

Under the assumption that the contests represented by the data in the win-loss matrix are independent contests, the weights $\{w_i\}$ can be estimated by maximum likelihood. A unique solution will exist for each set of comparable instructors and it can be found by iterative techniques. The log of the likelihood function is

$$\ell n\, L = \sum_{i \neq j}\sum a_{ij}[\ell n\, w_i - \ell n(w_i + w_j)] \qquad (5.7)$$

where $\{a_{ij}\}$ are the elements of the win-loss matrix. The system of partial derivatives of (5.7) with respect to w_r leads to the system of equations

$$\sum_j a_{rj} = w_r \sum_j N_{rj}/(w_r + w_j) \qquad (5.8)$$

which can be solved by iteration using the "win percentages" of the instructors to initialize the $\{w_r\}$. The quantities

$$\phi_i = - \sum_j a_{ij}[\ell n\, w_i - \ell n(w_i + w_j)] \qquad (5.9)$$

can be used for measuring the stress in estimating w_i. If the total stress is large then the likelihood function is flat near the maximum.

Both the Ford weights and the classical scaling models were applied to the 1979 data. Due to space limitations in the computer, the application was limited to a 60 by 60 win-loss matrix consisting of the top 60 ranked instructors of 1979 ranked according to the newly proposed score S^* (with $p = .75$ and weights 5:4:3:1). This class of instructors formed a

	ID	N	FW	Stress	ID	N	CS	Stress
1	2	51	1.33	3.96	2	51	0.45	6.41
2	1	71	1.15	2.80	1	71	0.40	2.40
3	8	23	1.10	2.06	8	23	0.35	2.98
4	3	57	0.97	3.43	3	57	0.30	4.03
5	4	67	0.85	3.96	4	67	0.26	3.55
6	7	72	0.84	1.92	7	72	0.25	1.27
7	11	72	0.79	1.67	11	72	0.21	0.62
8	16	44	0.75	3.26	16	44	0.20	2.76
9	12	47	0.70	2.28	12	47	0.17	1.87
10	23	30	0.64	1.24	23	30	0.13	0.88
11	5	83	0.62	3.42	5	83	0.12	4.74
12	21	26	0.61	1.20	21	26	0.10	0.69
13	28	20	0.61	0.89	17	28	0.09	1.08
14	17	28	0.60	1.26	13	50	0.09	1.25
15	13	50	0.59	2.10	33	25	0.09	0.68
16	33	25	0.59	1.23	28	20	0.08	0.87
17	35	47	0.57	2.16	55	20	0.08	0.74
18	55	20	0.56	0.75	35	47	0.07	1.52
19	34	38	0.55	1.34	18	23	0.07	0.58
20	18	23	0.55	0.89	34	38	0.06	1.41
21	52	45	0.55	0.86	52	45	0.04	1.20
22	39	33	0.53	1.23	39	33	0.04	0.68
23	25	45	0.52	1.62	25	45	0.04	1.03
24	19	21	0.51	0.53	19	21	0.03	0.55
25	47	41	0.51	1.48	15	96	0.03	2.18
26	15	96	0.51	3.00	26	56	0.03	0.95
27	26	56	0.51	1.53	47	41	0.02	0.89
28	50	35	0.51	1.22	36	39	0.02	1.05
29	36	39	0.51	1.71	50	35	0.02	1.07
30	6	128	0.49	3.47	6	128	0.00	1.82
31	49	44	0.48	1.72	49	44	¯0.01	0.79
32	45	83	0.47	1.88	32	48	¯0.01	0.97
33	32	48	0.47	1.48	45	83	¯0.02	3.16
34	51	50	0.46	1.79	54	45	¯0.02	0.83
35	27	90	0.46	3.66	44	50	¯0.03	0.95
36	54	45	0.45	1.14	51	50	¯0.03	1.21
37	44	50	0.45	1.17	27	90	¯0.03	2.19
38	24	78	0.45	1.57	24	78	¯0.03	1.83
39	48	33	0.44	1.26	48	33	¯0.04	0.87
40	30	96	0.44	3.33	30	96	¯0.04	2.12
41	53	51	0.44	1.54	53	51	¯0.05	1.16
42	10	83	0.41	2.00	10	83	¯0.07	1.23
43	56	48	0.41	1.15	56	48	¯0.08	1.57
44	22	53	0.40	1.26	22	53	¯0.08	1.14
45	38	72	0.40	1.82	38	72	¯0.09	1.88
46	58	32	0.39	0.65	58	32	¯0.10	0.93
47	41	39	0.38	0.71	41	39	¯0.11	1.06
48	31	81	0.37	1.94	29	41	¯0.11	0.66
49	29	41	0.37	0.85	31	81	¯0.12	1.96
50	14	94	0.33	1.55	9	36	¯0.13	1.44
51	9	36	0.33	0.74	14	94	¯0.16	1.23
52	20	37	0.33	0.63	20	37	¯0.17	0.39
53	46	80	0.32	1.52	46	80	¯0.18	2.91
54	42	61	0.31	1.03	42	61	¯0.20	1.43
55	59	37	0.30	0.66	57	36	¯0.20	1.46
56	57	36	0.30	0.56	59	37	¯0.21	0.78
57	60	112	0.30	1.81	60	112	¯0.21	4.42
58	43	85	0.30	1.49	43	85	¯0.22	3.61
59	40	69	0.20	1.13	40	69	¯0.25	2.18
60	37	26	0.12	0.06	37	26	¯0.51	1.85

LEGEND: FW - Weights of the Ford Model
CS = Score by Classical Scaling
ID = Instructor identification: rank by basic 1979 scores S*
Stress = % of total stress.

TABLE 5.1. Summary of Paired Comparisons Study

single set of comparable instructors. The results appear in Table 5.1. The two methods agree quite well within themselves (Kendall's coefficient of concordance is 0.98). The total stress is rather large for each (the standard deviation in (5.5) is 0.48). Each of the scalings in Table 5.1 are presented in top-to-bottom order. The individual instructors are identified by their ID numbers which are their ranks (top-to-bottom) using (4.3). See Appendix C for further summaries of the 1979 data.

Although the two scaling methods agree within themselves, they are in only modest agreement with the ranking provided by the score S^*. It was decided to vary the inputs (weights w and exponent p) of S^* to see if this condition could be improved. The results appear in Table 5.2 which contains the values of Kendall's tau.

Wts p	1.0	.75	.5
W0 = 3:2:1:0	.44/.45	.42/.43	.32/.33
W1 = 5:4:3:1	.49/.49	.41/.42	.27/.28
W2 = 4:3:2:1	.48/.48	.40/.41	.26/.27
W3 = 1:1:1:0	.48/.49	.43/.44	.29/.30
W4 = 1:1:1:1	.56/.56	.26/.25	.06/.07

Kendall's τ Computed Between S^* and FW/CS

TABLE 5.2. Concordance Coefficients

1979 Data

Examination of this table shows that the best agreement of S^* with paired comparison scalings occurs at the original normalization, $p = 1$. This in turn suggests that this scaling also has a bias favoring instructors with small values of N. On the other hand, the changes in instructor position provided by the paired comparison scoring provides information that is unavailable in S^*. When an instructor who is ranked low down according to S^* appears high on a paired comparison list, we know that his low value of S^* is due to the fact that he has been compared with a strong set of instructors on the ballots. Similarly when we see an instructor who is ranked high on the S^* list appear low down on a paired comparison list we know that he faced weaker competition on the ballots.

Since it is difficult to get the advantages of S^* and paired comparisons together in a single scoring system, it seems wise to include both in the data summaries to be reviewed by the selection committee. However, the Ford weights and the classical scaling are not both needed. The latter is selected arbitrarily for use in the proposed data summary system.

VI. CONCLUSIONS AND RECOMMENDATIONS

Generally, the award has a good reputation. The system has always selected a good teacher for the recipient. The practice of choosing as winner an instructor with a good "past performance" record in the balloting is a good one, but changes are needed to remove biases (with N) when considering past performance. The ballot validation restriction of $5 \leq K \leq 25$ is arbitrary but wise since it reduces the negative effect of a voter listing a large number of instructors. Further adjustment for this effect is possible (using (4.2) with $w_4 > 0$) and is considered worthwhile.

In order for the system to maintain its dignity the selection committee should monitor the responses in terms of watching the distributions of ballots returned by curricular area relative to the number sent out, the distributions of number of faculty listed on ballots by curricula area, and distribution of ballots to alumni by curricula area. Inbalances in these distributions should be avoided and investigated when they occur as they may signal the presence of unfair practices. Computer programs are being written to provide data summaries of these distributions.

Because of the arbitrary inputs, the ranking of instructors is arbitrary and it is nonsensical to make firm distinctions for the historical record. The balloting system is geared to identify good instructors. It is reasonable to select a "top 5%" and a "next 15%" for storage--the former being marked with an "A" and the latter with a "B". All other eligible faculty should be identified with an "E" and ineligible faculty with an "I". The values of N should be kept also. A set of IBM cards is being prepared to contain this record. Also a computer program is being written to facilitate the yearly updating of the record.

The exponent p and weights $W = (w_1:w_2:w_3:w_4)$ of the new score S^* are arbitrary inputs. The values $p = .75$ and $W = (5:4:3:1)$ were chosen by the 1978 Selection Committee for the following reasons:

a) The exponent $p = .75$ serves to avoid a correlation with N of the high scores.

b) The weights $W = (5:4:3:1)$ provide equally spaced weights for the ranking positions and also allows a positive contribution to the score of eligible faculty who are listed on a ballot but not ranked. The policy helps to offset the manipulative effect of ballots that list a large number K of instructors. See Equation (4.3).

Future Selection Committees should monitor the performance of the inputs p and W. Computer programs that produce output similar to that in Appendix A are being written for this purpose.

Finally, the data summaries supplied to the Selection Committee will consist of three pages of output:

(i) For each of the top 60 instructors (according to the score S^*) will be listed (see Appendix D)

Rank by decreasing values of S^*

N: the number of ballots that identify him

D: the number of ballots that provide a supporting statement

S^*: the score

X_1, X_2, X_3: the total accumulated 1st, 2nd and 3rd place votes

Z_1, Z_2, Z_3, Z_4: the total accumulated quantile counts of the four types of responses

(ii) For each string of comparable instructors in the top 60 will be listed (see Table (5.1))

Rank according to localized (and decreasing) CS

Rank (ID) according to S^*

N: the number of ballots that identify him

CS: score by classical scaling

Stress: as a percentage of total stress for that string.

(iii) For each of the top 30 instructors will be listed (similar to Appendix D)

Current values of N, S^*, Rank by CS and string identification

Recorded values of N and grade code (A,B,E,I) for each of the past seven years.

It is expected that the first page will provide the committee with adequate numerical information so that its members can visualize the entire body of data. This can serve as a general basis of comparison. The second page will help identify such instructors who have a relatively low S* because they found strong competition or who have a relatively high S* because they faced weaker competition. The past performance data is contained on page 3 together with excerpts from the first two pages. The information on these three pages provides the basis for designating a winner, the top 5% of the eligible instructors, and the next 15%.

APPENDIX A

Further study of the distribution of N (number of ballots that identify the instructor) as it relates to the large scores S^* ($S^* = (w_1 Z_1 + w_2 Z_2 + w_3 Z_3 + w_4 Z_4) \div N^p$).

To set the standard, the overall distribution of N for 1978 is summarized as follows.

MIN	LOQ	MEDIAN	UPQ	MAX	SIZE
1.00	15.50	28.00	42.00	101.00	212.00

where LOQ and UPQ stand for lower and upper quartiles, respectively. The above format is the output of an APL program called CONDENSE and is used liberally in what follows. See [9].

More specifically, a program called DECOMP was written which takes the top 60 scores S^* (for given weights W and power p), decomposes them into six sets containing 10 instructors each (according to decreasing S^*), and summarizes the corresponding six sets of 10 values of N. These summaries are in the form of CONDENSE and each is followed by a set of vertical BOXPLOTS (Ref. [9]). The following seven figures contain the results. The choice $p = .75$ is supported rather generally as a scoring scheme whose high scores favor neither large nor small values of N. It also appears that attention should be drawn to Fig. A-7, especially in the light of the behavior of the weight system $W3 = (1:1:1:0)$ coupled with $p = 1$ in Table 5.2.

10 *DECOMP N*

MIN	*LOQ*	*MEDIAN*	*UPQ*	*MAX*	*SIZE*
36.00	48.00	60.50	71.00	101.00	10.00
13.00	26.00	40.00	66.00	83.00	10.00
6.00	20.00	33.00	50.00	66.00	10.00
25.00	33.00	44.00	59.00	61.00	10.00
11.00	24.00	28.50	42.00	50.00	10.00
23.00	34.00	37.50	41.00	52.00	10.00

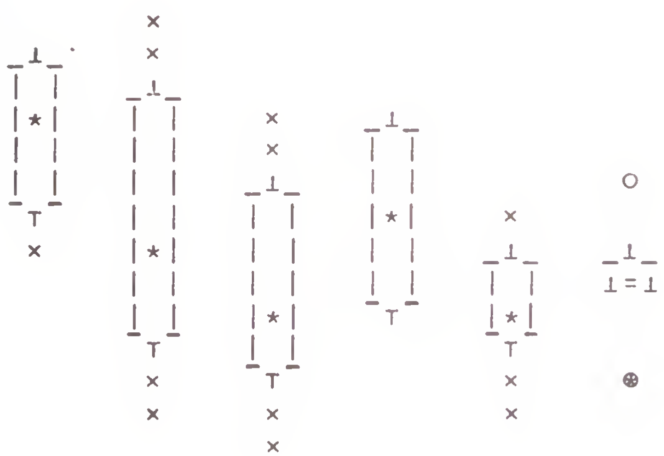


FIGURE A-1

$$S^{*} = (5Z_1 + 4Z_2 + 3Z_3 + Z_4) \div N^{.5}$$

10 *DECOMP N*

MIN	*LOQ*	*MEDIAN*	*UPQ*	*MAX*	*SIZE*
6.00	18.00	48.00	58.00	71.00	10.00
18.00	26.00	33.50	66.00	101.00	10.00
11.00	25.00	41.50	50.00	76.00	10.00
22.00	24.00	30.00	59.00	83.00	10.00
27.00	34.00	42.50	53.00	61.00	10.00
16.00	37.00	39.50	49.00	59.00	10.00

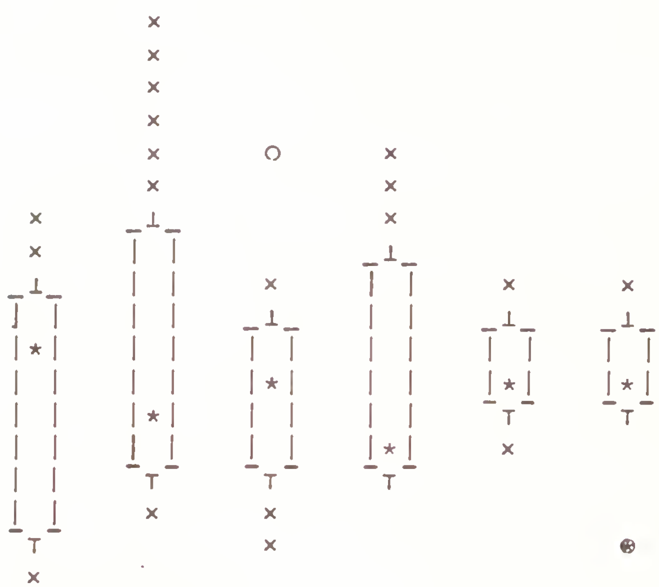


FIGURE A-2

$$S^{*} = (5Z_1 + 4Z_2 + 3Z_3 + Z_4) \div N^{.75}$$

10 DECOMP N

MIN	LOQ	MEDIAN	UPQ	MAX	SIZE
36.00	48.00	60.50	71.00	101.00	10.00
13.00	26.00	40.00	66.00	83.00	10.00
6.00	20.00	42.00	59.00	66.00	10.00
18.00	32.00	42.50	53.00	61.00	10.00
11.00	24.00	28.50	42.00	50.00	10.00
23.00	34.00	37.50	41.00	52.00	10.00

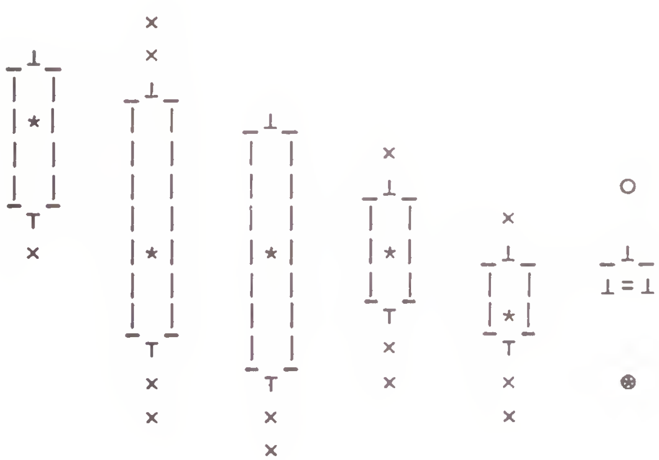


FIGURE A-3

$$S^* = (4Z_1 + 3Z_2 + 2Z_3 + Z_4) \div N^{.5}$$

10 *DECOMP N*

MIN	*LOQ*	*MEDIAN*	*UPQ*	*MAX*	*SIZE*
6.00	18.00	48.00	58.00	71.00	10.00
18.00	20.00	33.50	66.00	101.00	10.00
11.00	28.00	41.50	50.00	76.00	10.00
22.00	25.00	30.00	61.00	83.00	10.00
16.00	33.00	42.50	53.00	61.00	10.00
27.00	34.00	39.50	49.00	59.00	10.00

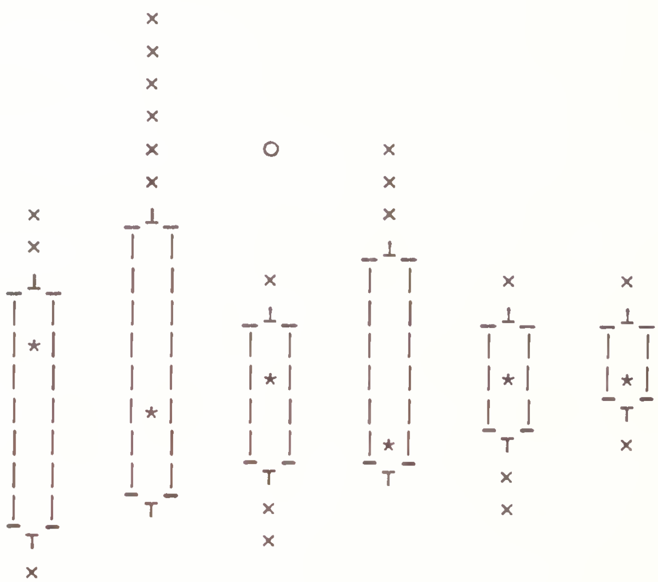


FIGURE A-4

$$S^* = (4Z_1 + 3Z_2 + 2Z_3 + Z_4) \div N^{.75}$$

10 *DECOMP N*

MIN	*LOQ*	*MEDIAN*	*UPQ*	*MAX*	*SIZE*
6.00	13.00	19.00	36.00	58.00	10.00
20.00	28.00	41.50	63.00	71.00	10.00
3.00	22.00	35.50	66.00	101.00	10.00
16.00	25.00	30.00	46.00	63.00	10.00
2.00	8.00	25.00	45.00	76.00	10.00
17.00	27.00	39.50	59.00	83.00	10.00

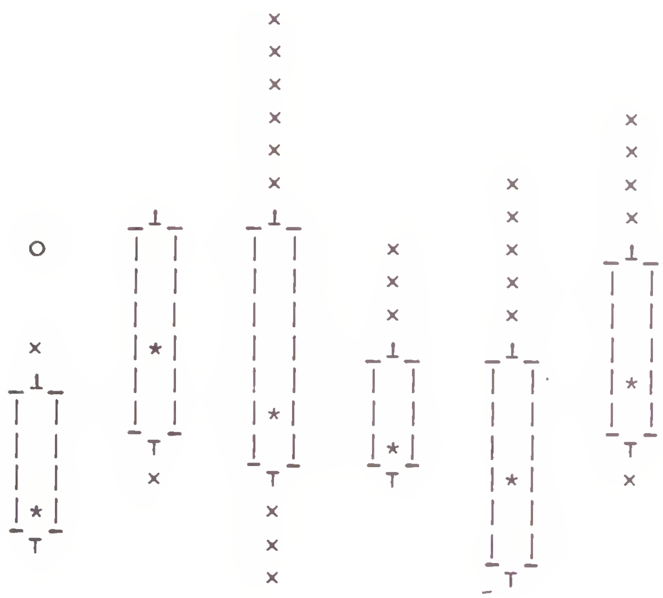


FIGURE A-5

$$S^{*} = (4Z_1 + 3Z_2 + 2Z_3 + Z_4) \div N$$

10 *DECOMP N*

MIN	*LOQ*	*MEDIAN*	*UPQ*	*MAX*	*SIZE*
36.00	48.00	60.50	71.00	101.00	10.00
13.00	18.00	31.50	66.00	83.00	10.00
6.00	25.00	42.00	59.00	66.00	10.00
20.00	30.00	33.50	59.00	61.00	10.00
11.00	23.00	39.00	45.00	50.00	10.00
24.00	27.00	37.50	46.00	52.00	10.00

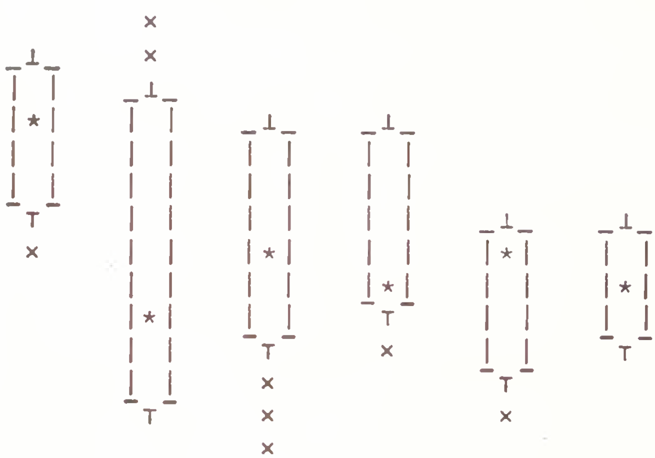


FIGURE A-6

$$S^* = (Z_1 + Z_2 + Z_3) \div N^{.5}$$

10 *DECOMP N*

MIN	*LOQ*	*MEDIAN*	*UPQ*	*MAX*	*SIZE*
6.00	18.00	23.00	48.00	71.00	10.00
2.00	20.00	30.00	48.00	63.00	10.00
7.00	25.00	35.50	50.00	78.00	10.00
3.00	24.00	40.00	63.00	101.00	10.00
16.00	27.00	35.00	66.00	83.00	10.00
11.00	21.00	41.50	45.00	50.00	10.00

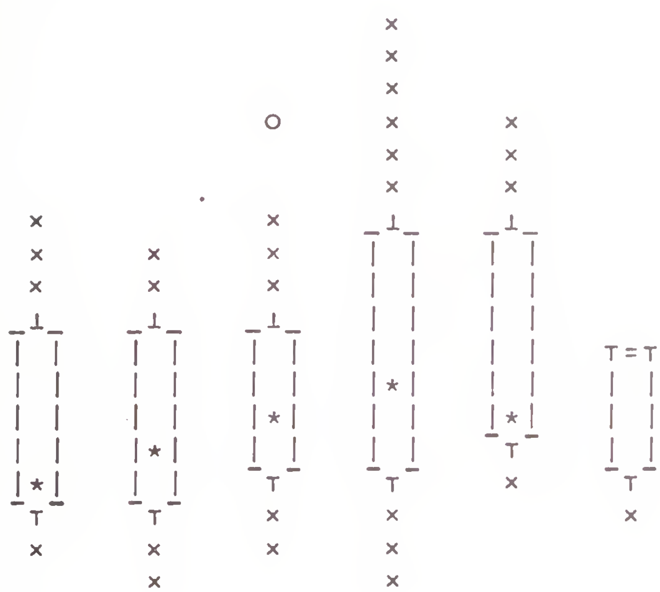


FIGURE A-7

$$S^{*} = (Z_1 + Z_2 + Z_3) \div N$$

APPENDIX B

EFFECT OF VARYING INPUTS

The 1978 data were used to explore the effect of varying the exponent p and the weighting system W. Specifically,

$$p = .75, .5$$

and

$$W0 = 3:2:1:0$$

$$W1 = 5:4:3:1$$

$$W2 = 4:3:2:1$$

$$W3 = 1:1:1:0$$

$$W4 = 1:1:1:1$$

The index numbers in Table B-1 are the ranks according to the original score S with $p = 1$ and $W = 4:2:1$. Table B-2 contains the corresponding values of N.

Rank	p = .75 W0	p = .75 W1	p = .75 W2	p = .75 W3	p = .75 W4	p = .5 W0	p = .5 W1	p = .5 W2	p = .5 W3	p = .5 W4
1	4	4	4	4	4	4	4	4	4	28
2	2	2	2	5	10	10	10	10	10	4
3	1	5	5	10	7	5	7	18	5	10
4	5	10	10	2	5	7	5	7	7	18
5	3	1	1	1	18	11	18	5	18	32
6	10	7	7	7	21	18	11	11	11	21
7	7	3	3	3	28	28	28	28	21	7
8	11	11	11	11	11	2	21	21	28	11
9	18	18	18	8	32	15	15	15	32	36
10	15	15	15	13	15	21	32	32	15	60
11	8	21	21	18	2	3	24	24	8	24
12	12	8	8	15	36	32	2	36	2	31
13	16	16	12	21	24	24	36	2	33	15
14	21	12	28	16	31	12	3	31	16	5
15	9	28	16	19	33	16	31	3	3	48
16	28	13	32	32	8	8	16	16	24	33
17	13	32	9	33	16	1	12	12	13	57
18	6	9	24	9	3	36	8	8	19	55
19	14	24	13	28	12	31	33	33	36	37
20	24	19	14	12	60	27	60	60	60	71
21	19	14	19	30	37	33	27	27	31	76
22	32	33	31	14	23	9	48	48	27	61
23	31	31	33	24	48	19	19	23	12	16
24	33	36	36	27	13	23	23	37	55	27
25	23	23	23	6	57	14	37	19	30	79
26	27	6	6	34	19	13	1	1	1	126
27	36	27	27	23	55	48	9	9	37	23
28	20	30	37	31	34	60	13	57	9	12
29	30	37	30	37	30	37	55	55	23	85
30	37	34	34	26	1	6	14	14	48	127
31	26	48	48	36	27	26	57	13	34	135
32	25	60	60	55	14	34	34	34	14	53
33	34	26	20	60	79	25	30	53	26	8
34	22	20	26	40	9	30	61	61	61	91
35	48	55	25	48	61	57	26	71	40	141
36	35	40	22	70	76	55	53	30	70	107
37	40	25	35	20	53	20	71	45	76	123
38	39	35	39	42	71	35	76	26	57	115
39	60	22	40	35	85	53	44	44	71	34
40	42	42	57	61	45	22	45	76	35	45
41	53	39	53	25	44	45	25	25	53	118
42	29	53	55	39	67	44	35	35	79	44
43	44	57	42	22	39	61	79	79	44	159
44	45	44	44	67	40	39	56	56	6	129
45	57	61	45	53	91	40	40	20	85	19
46	55	45	61	52	77	71	20	39	67	77
47	52	52	56	44	70	56	39	22	42	67
48	56	56	52	77	42	42	85	85	77	56
49	61	71	71	76	107	52	6	40	20	104
50	54	70	29	79	56	76	22	6	25	82
51	71	76	76	57	127	54	52	52	52	101
52	41	67	79	41	35	79	67	67	56	145
53	58	79	67	85	52	29	42	42	39	3
54	67	29	54	71	104	67	70	54	22	30
55	38	85	70	56	126	58	77	70	45	70
56	76	54	85	69	118	63	54	77	104	63
57	70	77	58	29	54	85	91	63	69	2
58	47	58	77	45	73	70	63	91	82	54
59	17	63	63	94	25	77	82	82	94	89
60	51	73	51	81	26	51	73	73	91	73

TABLE B-1. Index Numbers for Several Scoring Systems

	p = .75					p = .5				
Rank	W0	W1	W2	W3	W4	W0	W1	W2	W3	W4
1	58	58	58	58	58	58	58	58	58	101
2	13	13	13	36	71	71	71	71	71	58
3	6	36	36	71	48	36	48	68	36	71
4	36	71	71	13	36	48	36	48	48	68
5	18	6	6	6	68	57	68	36	68	78
6	71	48	48	48	63	68	57	57	57	63
7	48	18	18	18	101	101	101	101	63	48
8	57	57	57	57	57	13	63	63	101	57
9	68	68	68	26	78	48	48	48	78	76
10	48	48	48	18	48	63	78	78	48	83
11	26	63	63	68	13	18	66	66	26	66
12	32	26	26	48	76	78	13	76	13	63
13	35	35	32	63	66	66	76	13	45	48
14	63	32	101	35	63	32	18	63	35	36
15	20	101	35	28	45	35	63	18	18	66
16	101	18	78	78	26	26	35	35	66	45
17	18	78	20	45	35	6	32	32	18	61
18	11	20	66	20	18	76	26	26	28	59
19	20	66	18	101	32	63	45	45	76	46
20	66	28	20	32	83	50	83	83	83	61
21	28	20	28	25	46	45	50	50	63	59
22	78	45	63	20	38	20	66	66	50	53
23	63	63	45	66	66	28	28	38	32	35
24	45	76	76	50	18	38	38	46	59	50
25	38	38	38	11	61	20	46	28	25	50
26	50	11	11	33	28	18	6	6	6	75
27	76	50	50	38	59	66	20	20	46	38
28	22	25	46	63	33	83	18	61	20	32
29	25	46	25	46	25	46	59	59	38	49
30	46	33	33	32	6	11	20	20	66	64
31	32	66	66	76	50	32	61	18	33	74
32	30	83	83	59	20	33	33	33	20	45
33	33	32	22	83	50	30	25	45	32	26
34	24	22	32	25	20	25	53	53	53	52
35	66	59	30	66	53	61	32	61	25	73
36	30	25	24	34	59	59	45	25	34	52
37	25	30	30	22	45	22	61	43	59	70
38	27	30	27	23	61	30	59	32	61	65
39	83	24	25	30	49	45	42	42	61	33
40	23	23	61	53	43	24	43	59	30	43
41	45	27	45	30	42	43	30	30	45	56
42	16	45	59	27	37	42	30	30	50	42
43	42	61	23	24	27	53	50	50	42	59
44	43	42	42	37	25	27	42	42	11	65
45	61	53	43	45	52	25	25	22	49	28
46	59	43	53	33	41	61	22	27	37	41
47	33	33	42	42	34	42	27	24	23	37
48	42	42	33	41	23	23	49	49	41	42
49	53	61	61	59	52	33	11	25	22	46
50	38	34	16	50	42	59	24	11	30	47
51	61	59	59	61	64	38	33	33	33	51
52	7	37	50	7	30	50	37	37	42	58
53	31	50	37	49	33	16	23	23	27	18
54	37	16	38	61	46	37	34	38	24	25
55	11	49	34	42	75	31	41	34	43	34
56	59	38	49	25	56	41	38	41	46	41
57	34	41	31	16	38	49	52	41	25	13
58	17	31	41	43	37	34	41	52	47	38
59	3	41	41	30	30	41	47	47	30	41
60	27	37	27	2	32	27	37	37	52	37

TABLE B-2. Values of N Corresponding to Table B-1.

APPENDIX C

DATA SUMMARY OF TOP SIXTY SCORES BY CURRENT METHOD AND BY THE PROPOSED NEW METHOD

Rank	S	D	N	X1	X2	X3
1	3.33	3	6	4	2	0
2	2.69	6	13	8	1	1
3	2.22	6	18	8	4	0
4	2.00	17	58	24	9	2
5	1.92	9	36	12	8	5
6	1.73	4	11	4	1	1
7	1.58	9	48	14	7	6
8	1.58	5	26	7	4	5
9	1.55	4	20	5	5	1
10	1.54	12	71	18	15	7
11	1.51	9	57	16	9	4
12	1.50	6	32	11	1	2
13	1.50	2	18	3	5	5
14	1.40	4	20	5	3	2
15	1.38	10	48	13	4	6
16	1.37	4	35	7	9	2
17	1.33	1	3	1	0	0
18	1.32	4	68	16	11	4
19	1.25	5	28	5	5	5
20	1.14	4	22	5	2	1
21	1.13	6	63	10	12	7
22	1.08	4	24	5	3	0
23	1.05	5	38	7	5	2
24	1.05	6	66	13	7	3
25	1.03	5	30	7	1	1
26	1.03	2	32	5	4	5
27	1.02	6	50	8	7	5
28	1.01	13	101	18	10	10
29	1.00	2	16	3	2	0
30	1.00	2	25	2	6	5
31	0.97	7	63	11	7	3
32	0.96	4	78	9	16	7
33	0.96	4	45	4	11	5
34	0.94	3	33	4	5	5
35	0.93	3	30	5	3	2
36	0.92	10	76	15	2	6
37	0.91	5	46	6	6	6
38	0.91	2	11	2	1	0
39	0.89	4	27	5	1	2
40	0.88	0	25	3	3	4
41	0.86	0	7	0	3	0
42	0.83	1	23	2	5	1
43	0.80	1	5	1	0	0
44	0.79	3	42	6	3	3
45	0.77	4	43	7	2	1
46	0.76	1	17	3	0	1
47	0.76	2	17	3	0	1
48	0.76	5	66	6	12	2
49	0.75	1	8	1	1	0
50	0.75	3	16	3	0	0
51	0.74	3	27	5	0	0
52	0.73	2	33	3	5	2
53	0.71	4	45	4	8	0
54	0.71	5	38	6	1	1
55	0.69	4	59	6	2	13
56	0.69	3	42	5	3	3
57	0.69	3	61	8	3	4
58	0.68	2	31	4	2	1
59	0.67	4	24	4	0	0
60	0.66	4	83	6	11	9

TABLE C-1. Score Summary 1978. (Old System W = 4,2,1; p = 1).

Rank	N	D	S*	Z1	Z2	Z3	Z4	X1	X2	X3
1	58	17	7.18	21.61	7.26	1.63	8.90	24	9	2
2	13	6	6.56	7.60	0.90	0.70	1.21	8	1	1
3	36	9	6.49	10.96	6.70	3.32	3.78	12	8	5
4	71	12	6.34	10.30	11.04	5.32	9.92	18	15	7
5	6	3	6.26	3.63	1.49	0.00	0.00	4	2	0
6	46	9	6.06	12.89	5.76	5.12	7.66	14	7	6
7	18	6	5.88	7.11	3.42	0.00	2.14	8	4	0
8	57	9	5.66	14.22	7.49	2.24	9.71	16	9	4
9	68	4	5.52	14.47	9.13	2.85	13.27	16	11	4
10	48	10	5.32	12.02	3.58	4.62	8.65	13	4	6
11	63	6	5.14	8.93	10.33	5.39	12.71	10	12	7
12	26	5	5.13	6.32	3.07	3.86	3.59	7	4	5
13	35	4	4.88	6.23	7.21	1.38	6.10	7	9	2
14	32	6	4.87	10.16	0.82	1.65	6.54	11	1	2
15	101	13	4.83	16.16	7.69	6.76	22.08	18	10	10
16	18	2	4.71	2.58	3.98	3.49	1.84	3	5	5
17	78	4	4.67	8.08	12.46	5.55	15.71	9	16	7
18	20	4	4.59	4.30	4.18	0.63	3.34	5	5	1
19	66	6	4.49	11.68	6.13	2.20	14.48	13	7	3
20	28	5	4.46	4.52	4.07	3.77	4.05	5	5	5
21	20	4	4.41	4.60	2.51	1.67	3.71	5	3	2
22	45	4	4.38	3.67	9,15	4.17	8.67	4	11	5
23	63	7	4.23	9.97	5.80	2.08	15.25	11	7	3
24	76	10	4.15	13.62	1.64	4.47	18.82	15	2	6
25	38	5	4.12	6.27	4.51	1.63	8.72	7	5	2
26	11	4	4.10	3.67	0.83	0.50	1.59	4	1	1
27	50	6	4.05	7.16	5.38	3.29	9.02	8	7	5
28	25	2	3.97	1.77	4.99	3.69	4.47	2	6	5
29	46	5	3.86	5.34	4.73	4.01	10.55	6	6	6
30	33	3	3.81	3.55	4.31	3.43	7.10	4	5	5
31	66	5	3.66	5.42	9.58	1.20	15.67	6	12	2
32	83	4	3.62	5.51	8.55	6.51	18.39	6	11	9
33	32	2	3.61	4.34	3.24	2.96	5.00	5	4	5
34	22	4	3.55	4.54	1.60	0.73	4.80	5	2	1
35	59	4	3.45	5.56	1.42	9.23	12.33	6	2	13
36	25	0	3.45	2.81	2.47	3.05	5.43	3	3	4
37	30	5	3.44	6.39	0.67	0.83	6.91	7	1	1
38	30	3	3.41	4.60	2.33	1.56	6.68	5	3	2
39	24	4	3.36	4.55	2.05	0.00	5.52	5	3	0
40	23	1	3.33	1.99	4.33	0.82	5.78	2	5	1
41	27	4	3.32	4.64	0.86	1.74	7.43	5	1	2
42	45	4	3.29	3.75	6.59	0.00	12.05	4	8	0
43	61	3	3.28	7.38	2.25	2.86	17.04	8	3	4
44	42	3	3.23	5.27	2.53	2.00	10.79	6	3	3
45	53	1	3.21	2.78	6.50	3.33	13.15	3	8	5
46	43	4	3.16	6.43	1.52	0.79	12.18	7	2	1
47	33	2	3.07	2.70	4.26	1.23	8.07	3	5	2
48	42	3	3.04	4.59	2.42	2.12	10.64	5	3	3
49	61	2	3.01	3.59	5.89	2.82	15.79	4	7	4
50	34	0	3.01	0.00	5.82	3.83	7.56	0	7	5
51	59	1	2.98	2.82	5.70	3.78	15.28	3	7	5
52	37	1	2.97	1.91	4.13	2.95	9.70	2	5	4
53	50	1	2.90	0.95	7.40	2.46	13.99	1	9	3
54	16	2	2.96	2.73	1.53	0.00	3.96	3	2	0
55	49	0	2.86	1.88	4.37	4.27	13.16	2	5	5
56	38	5	2.80	5.18	0.88	0.83	10.89	6	1	1
57	41	0	2.79	0.91	4.81	3.67	10.43	1	6	5
58	31	2	2.70	3.66	1.55	0.97	8.32	4	2	1
59	41	0	2.65	3.50	2.90	0.85	11.33	4	4	1
60	37	1	2,62	2.78	2.62	1.46	10.55	3	3	2

TABLE C-2. Data Summary 1978. W = (5,4,3,1); p = .75.

APPENDIX D

DATA SUMMARIES FOR
THE 1979 TEACHING AWARD BALLOTING DATA

SCORING SUMMARY 1979
(weight = 5:4:3:1 and p = .75)

RANK	N	D	S	Z_1	Z_2	Z_3	Z_4	X_1	X_2	X_3
1	71	8	7.66	18.83	14.91	8.10	9.33	20	17	10
2	51	21	7.64	24.02	3.24	2.21	6.12	26	4	3
3	57	17	6.98	19.16	6.59	4.82	8.18	21	8	6
4	67	13	6.91	17.95	11.66	5.06	10.30	19	14	6
5	83	17	6.90	24.18	10.39	4.13	14.96	26	12	6
6	120	23	6.54	28.86	11.22	12.28	22.84	33	14	17
7	72	6	6.49	10.38	14.08	13.81	10.81	11	16	17
8	23	8	6.48	10.90	1.70	1.39	2.57	12	2	2
9	36	14	6.40	14.56	3.66	0.63	4.69	17	5	1
10	83	14	6.27	16.30	13.95	6.99	14.10	18	17	9
11	72	9	6.17	17.75	6.90	7.33	14.24	19	8	9
12	47	6	5.64	9.16	7.99	5.15	8.05	10	9	7
13	50	8	5.57	9.21	9.15	4.42	8.78	10	11	6
14	94	11	5.54	13.54	14.36	7.56	16.98	15	18	10
15	96	13	5.51	18.92	8.45	6.70	20.45	21	19	9
16	44	8	5.42	10.36	5.08	3.96	8.56	11	6	5
17	28	4	5.25	5.65	7.17	0.82	4.48	6	9	1
18	23	4	5.22	7.37	2.51	1.45	3.58	8	3	2
19	21	5	5.15	6.09	2.88	2.05	2.44	7	4	3
20	37	6	5.04	7.11	5.01	4.92	5.10	8	6	7
21	26	6	4.99	8.11	2.49	0.67	5.01	9	3	1
22	53	9	4.92	11.91	4.12	3.15	11.17	13	5	4
23	30	6	4.84	9.28	1.33	1.61	5.54	10	2	2
24	78	10	4.61	14.14	6.31	4.71	16.17	16	8	7
25	45	4	4.80	6.41	5.66	6.63	8.78	7	7	8
26	56	8	4.75	10.87	6.25	2.54	10.17	12	8	4
27	90	7	4.71	12.49	6.78	6.38	20.97	14	10	8
28	20	1	4.63	4.50	2.67	2.44	3.31	5	3	3
29	41	3	4.58	3.52	7.28	7.09	6.22	4	9	10
30	96	6	4.45	9.17	8.26	11.54	22.81	10	16	14
31	81	4	4.33	5.66	11.59	7.59	19.52	6	13	10
32	48	6	4.32	8.01	5.05	2.36	11.51	9	6	3
33	25	2	4.26	4.59	4.29	0.82	5.05	5	5	1
34	38	7	4.19	7.33	3.39	2.21	7.29	8	4	3
35	47	3	4.10	5.46	6.25	3.30	11.40	6	7	4
36	39	7	4.07	7.44	1.80	3.23	9.41	8	2	4
37	26	1	4.00	1.67	7.74	1.07	3.53	2	11	2
38	72	3	3.98	6.56	8.70	4.07	18.47	7	10	5
39	33	4	3.84	5.49	3.61	0.86	8.41	6	4	1
40	69	9	3.84	12.02	2.60	0.70	19.34	13	3	1
41	39	2	3.82	4.61	5.40	2.52	7.44	5	7	3
42	61	4	3.75	4.54	9.99	1.94	13.49	5	13	3
43	85	4	3.72	8.88	7.56	2.18	22.91	10	10	3
44	50	2	3.57	2.68	5.97	6.37	10.73	3	8	9
45	83	4	3.56	4.74	8.76	5.05	24.06	5	19	6
46	80	10	3.54	10.06	4.06	1.71	23.07	11	5	2
47	41	2	3.47	4.62	1.79	5.23	10.19	5	2	7
48	33	2	3.38	1.85	6.04	1.63	8.28	2	7	2
49	44	1	3.38	4.70	3.53	2.74	11.95	5	4	4
50	35	1	3.36	2.84	4.42	2.46	9.14	3	5	3
51	50	5	3.30	5.40	1.69	4.78	14.02	6	2	6
52	45	2	3.30	3.61	3.47	3.32	12.66	4	4	5
53	51	3	3.26	3.75	4.33	4.07	13.89	4	5	5
54	45	3	3.24	4.65	3.95	1.61	12.49	5	5	2
55	20	2	3.23	1.88	2.55	2.20	4.34	2	3	3
56	48	2	3.22	2.82	5.50	3.28	12.75	3	7	4
57	36	5	3.17	5.57	0.88	1.85	9.70	6	1	3
58	32	3	3.17	3.61	3.33	0.80	8.84	4	4	1
59	37	5	3.14	4.41	2.45	2.06	9.03	5	3	3
60	112	4	3.11	5.56	9.09	3.18	33.32	6	11	4

LEGEND: N = number of ballots that identify the instructor

D = number of ballots that contain a supporting statement

$S^* = \text{Score: } (5Z_1 + 4Z_2 + 3Z_3 + Z_4)/N^{.75}$

PREVIOUS PERFORMANCE DATA

RANK	1979			1978		1977 RANK	1976 RANK	1975 RANK	1974 RANK	1973 RANK	1972 RANK
	N	D	S*	N	Code						
1	71	8	7.66	0							
2	51	21	7.64	32	B		14	5	8		
3	57	17	6.98	0	I	11	10				8
4	67	13	6.91	68	A		7				
5	83	17	6.90	71	A						
6	128	23	6.54	63	B					9	
7	72	6	6.49	0	I					15	
8	23	8	6.48	6	A	5					
9	36	14	6.40	12	E						
10	83	14	6.27	0	I						
11	72	9	6.17	0	I	7			13	6	11
12	47	6	5.64	20	B						
13	50	8	5.57	35	B	15					
14	94	11	5.54	63	B						
15	96	13	5.41	66	B		11				
16	44	8	5.42	0	I						
17	28	4	5.25	16	E						

LEGEND: A = top 5%, B = next 15%, E = Eligible, I = Ineligible.

REFERENCES

[1] Black, Duncan. The Theory of Committees and Elections, Cambridge (1963).

[2] Capps, J.H. The Second Annual Award for Excellence in Teaching. A Study of Procedure and Analysis of Data. M.Sc. Thesis, Naval Postgraduate School (1971).

[3] DiMauro, P.V. The Fourth Annual Award for Excellence in Teaching. An Analysis of Procedure and Data. M.Sc. Thesis, Naval Postgraduate School (1973).

[4] Dudewicz, E. J. and Lee, Y.J. "How to Select the Best Contender," Annual Technical Conference Transactions of the American Socieity for Quality Control, 32 (1978).

[5] Ford, L.R., Jr., "Solution of a Ranking Problem From Binary Comparisons," American Mathematical Monthly, 64 (Oct. 1957).

[6] Freeman, M.F. and Tukey, J.W. "Transformations Related to the Angular and the Square Root." Annals of Math. Statistics, 21, 607-611 (1950).

[7] Geary, R. W. The First Annual Award for Excellence in Teaching. A Study of Procedure and Analysis of Data. M.Sc. Thesis, Naval Postgraduate School (1970).

[8] Gulliksen, H. "A Least Squares Solution for Paired Comparisons with Incomplete Data," Psychometika, 21, No. 2 (June 1956).

[9] McNeil, D.R. Interactive Data Analysis, Wiley (1977).

[10] Mosteller, F. and Tukey, J.W. "Data Analysis, Including Statistics," Chapter 2 of The Handbook of Social Psychology, Vol. 2, ed. by G. Lindzey and E. Aronson, Addison-Wesley (1968).

[11] Scango, P.D. The Third Annual Award for Excellence in Teaching. An Analysis of Procedure and Data. M.Sc. Thesis, Naval Postgraduate School (1972).

DISTRIBUTION LIST

	No. of Copies
Defense Documentation Center Cameron Station Alexandria, VA 22314	2
Library Code 0212 Naval Postgraduate School Monterey, CA 93940	2
Library Code 55 Naval Postgraduate School Monterey, CA 93940	1
Dean of Research Code 012 Naval Postgraduate School Monterey, CA 93940	1
Prof. E. J. Dudewicz Department of Statistics Ohio State University Columbus, Ohio 43210	1
Dr. Y. J. Lee Biostatistics and Data Management National Cancer Institute Bethesda, MD 20014	1
Prof. R. E. Bechhofer Mathematics Department Cornell University Ithaca, New York 14850	1
Prof. I. Olkin Statistics Department Stanford University Stanford, CA 94305	1
LT Harry Bishop Naval Submarine School New London, CT 06320	1

	No. of Copies
Prof. J. Gibbons Statistics Department Alabama University Tuscaloosa, AL 35401	1
Prof. Shanti Gupta Statistics Department Purdue University Lafayette, IN 47907	1
Naval Postgraduate School Monterey, CA 93940	
Attn: J. R. Borsting, Code 01	1
M. Woods, Code Code 011	1
D. A. Schrady, Code 013	1
T. Jayachandran, Code 53	1
J. Senger, Code 54	1
A. Andrus, Code 55	1
D. Barr, Code 55	1
J. Esary, Code 55	1
D. Gaver, Code 55	1
J. Hartman, Code 55	1
P. Jacobs, Code 55	1
H. Larson, Code 55	1
P.A.W. Lewis, Code 55	1
G. Lindsay, Code 55	1
P. Milch, Code 55	1
G. Poock, Code 55	1
R. R. Read, Code 55	20
R. Richards, Code 55	1
B. Shubert, Code 55	1
A. Washburn, Code 55	1
P. Zehna, Code 55	1
R. Stampfel, Code 55	1
A. Coppens, Code 61	1
J. Dyer, Code 61	1
R. Kelly, Code 61	1
A. Gerba, Code 62	1
E. Haderlie, Code 68	1
M. Kelleher, Code 69	1
Chairman, Code 34	1
Chairman, Code 52	1
Chairman, Code 53	1
Chairman, Code 54	1
Chairman, Code 55	1
Chairman, Code 56	1

	No. of Copies
Naval Postgraduate School Monterey, Ca. 93940	
Attn: Chairman, Code 61	1
Chairman, Code 62	1
Chairman, Code 63	1
Chairman, Code 67	1
Chairman, Code 68	1
Chairman, Code 69	1
Chairman, Code 71	1
Chairman, Code 73	1
Chairman, Code 74	1
DRMEC, J. Dawson, Code 64	1